Risk of Thyroid Cancer Based on Thyroid Ultrasound Imaging Characteristics
Results of a Population-Based Study

Rebecca Smith-Bindman, MD; Paulette Lebda, MD; Vickie A. Feldstein, MD; Dorra Sellami, MD; Ruth B. Goldstein, MD; Natasha Brasic, MD; Chengshi Jin, PhD; John Kornak, PhD

**IMPORTANCE** There is wide variation in the management of thyroid nodules identified on ultrasound imaging.

**OBJECTIVE** To quantify the risk of thyroid cancer associated with thyroid nodules based on ultrasound imaging characteristics.

**METHODS** Retrospective case-control study of patients who underwent thyroid ultrasound imaging from January 1, 2000, through March 30, 2005. Thyroid cancers were identified through linkage with the California Cancer Registry.

**RESULTS** A total of 8806 patients underwent 11 618 thyroid ultrasound examinations during the study period, including 105 subsequently diagnosed as having thyroid cancer. Thyroid nodules were common in patients diagnosed as having cancer (96.9%) and patients not diagnosed as having thyroid cancer (56.4%). Three ultrasound nodule characteristics—microcalcifications (odds ratio [OR], 8.1; 95% CI, 3.8-17.3), size greater than 2 cm (OR, 3.6; 95% CI, 1.7-7.6), and an entirely solid composition (OR, 4.0; 95% CI, 1.7-9.2)—were the only findings associated with the risk of thyroid cancer. If 1 characteristic is used as an indication for biopsy, most cases of thyroid cancer would be detected (sensitivity, 0.88; 95% CI, 0.80-0.94), with a high false-positive rate (0.44; 95% CI, 0.43-0.45) and a low positive likelihood ratio (2.0; 95% CI, 1.8-2.2), and 56 biopsies will be performed per cancer diagnosed. If 2 characteristics were required for biopsy, the sensitivity and false-positive rates would be lower (sensitivity, 0.52; 95% CI, 0.42-0.62; false-positive rate, 0.07; 95% CI, 0.07-0.08), the positive likelihood ratio would be higher (7.1; 95% CI, 6.2-8.2), and only 16 biopsies will be performed per cancer diagnosed. Compared with performing biopsy of all thyroid nodules larger than 5 mm, adoption of this more stringent rule requiring 2 abnormal nodule characteristics to prompt biopsy would reduce unnecessary biopsies by 90% while maintaining a low risk of cancer (5 per 1000 patients for whom biopsy is deferred).

**CONCLUSIONS AND RELEVANCE** Thyroid ultrasound imaging could be used to identify patients who have a low risk of cancer for whom biopsy could be deferred. On the basis of these results, these findings should be validated in a large prospective cohort.

Published online August 26, 2013.

**Author Affiliations:** Department of Radiology and Biomedical Imaging, University of California, San Francisco (Smith-Bindman, Feldstein, Goldstein, Brasic); Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco (Smith-Bindman, Jin, Kornak); Department of Epidemiology and Biostatistics, University of California, San Francisco (Smith-Bindman, Feldstein, Goldstein); Department of Radiology, Cleveland Clinic, Cleveland, Ohio (Lebda); California Advanced Imaging Medical Associates, Mills-Peninsula Hospital, San Mateo (Sellami).

**Corresponding Author:** Rebecca Smith-Bindman, MD, Department of Radiology and Biomedical Imaging, University of California, 350 Parnassus Ave, Ste 307, San Francisco, CA 94143-0336 (rebecca.smith-bindman@ucsf.edu).
Ultrasound imaging has replaced nuclear medicine as the most frequently used imaging test of the thyroid. The increase in the use of thyroid ultrasound imaging by radiologists, endocrinologists, and head and neck surgeons has led to the discovery of large numbers of asymptomatic thyroid nodules, which may occur in 50% or more of adults, as well as a rapid increase in the diagnosis of thyroid cancer. In contrast, clinically apparent thyroid cancer is rare, affecting 1 in 10,000 people annually and less than 1% of individuals throughout their lives. Because of the high prevalence of nodules and the rarity of symptomatic cancer, only a small percentage of thyroid nodules are malignant. Uncertainty about which nodules may harbor cancer and lack of evidence-based management guidelines have resulted in a myriad of conflicting recommendations regarding which nodules warrant biopsy, frequent thyroid biopsies, and the overdiagnosis of thyroid cancers that would otherwise likely have remained asymptomatic in the absence of detection.

Although many studies have analyzed the association between the ultrasound imaging characteristics of thyroid nodules and the risk of thyroid cancer, most studies are small, and all limited their analysis to patients who underwent biopsy, in which the decision to biopsy was influenced by the ultrasound imaging result. This ascertainment bias will overestimate the risk of cancer associated with thyroid biopsy and the accuracy of ultrasound imaging. The information that is most important to patients and physicians managing care is quantifying the risk of cancer associated with a nodule with a particular imaging characteristic, and no prior publication can accurately provide this information. This obstacle has hindered the development of an evidence-based strategy for determining which nodules should be biopsied because of an elevated cancer risk. The purposes of this study were to determine the ultrasound imaging characteristics that are associated with cancer and to use this information to create a standardized system for interpreting thyroid ultrasound imaging.

Methods

We conducted a retrospective, case-control study at the University of California, San Francisco, of consecutive patients who underwent thyroid ultrasound imaging from January 1, 2000, through March 30, 2005. A waiver of patient informed consent was obtained. This study was approved by the institutional review board of the University of California, San Francisco. Patients were excluded if they had a prior unilateral or bilateral thyroidectomy for benign or malignant disease.

Linkage to Population Cancer Registry

Cancers identified in the cohort were identified through linkage with the California Cancer Registry, a population-based cancer registry collecting cancer incidence and mortality data for all of California. The registry is a collaboration between the Cancer Surveillance Section of the California Department of Public Health, The Public Health Institute, and 8 regional cancer registries that by legislative mandate have collected cancer incidence data from hospitals and other facilities across the state since 1988. The registry is certified by the North American Association of Central Cancer Registries as meeting their highest standard for completeness of cancer ascertainment, reflecting capture of more than 97% of cancers diagnosed in the state. We included thyroid cancers diagnosed through March 30, 2007, allowing a minimum 2 years of follow-up after the last enrolled patient’s ultrasound imaging during which a cancer could be diagnosed and at least 2 years of further follow-up to ensure reporting to the registry. Patients diagnosed as having nonthyroid malignant neoplasms (other than skin cancer) were excluded to prevent the inclusion of the rare but theoretically possible metastatic cancer to the thyroid because these metastatic cancers would not be captured by the cancer registry. All patients diagnosed as having cancer (cases, and a sample of patients not diagnosed as having cancer (controls) matched 4 to 1 to the cancer patients for age, sex, and year of the ultrasound examination were selected for detailed review of the sonogram.

Characterizations of the Ultrasound Examinations

We retrieved and reviewed the results of the ultrasound examinations of 96 cancer patients (91.4%) from the Radiology Picture Archiving and Communication System and 369 controls. Each ultrasound examination result was reviewed independently by 2 board-certified radiologists (R.S.-B., P.L., V.A.F., D.S., R.B.G., and N.B.) masked to cancer status. Disagreement was resolved by consensus. For each patient, each reviewer independently recorded the number, size, and characteristics of all nodules larger than 5 mm. There was good to outstanding agreement (κ = 0.73 to 1.0) in the categorization of the specific ultrasound imaging characteristics.

Assignment of Cancer Status to Thyroid Nodules

In patients selected as controls, all nodules were considered benign. In 43 cancer patients (44.9%), a single nodule was identified and was considered malignant. In 50 cancer patients (52.1%), multiple thyroid nodules were identified. To ensure correct attribution of cancer to the correct nodule, one of the authors (D.S.) was unmasked, and patient records (radiology, pathology, and surgery) were reviewed to determine which nodules were malignant. In the small number of cases for which

<table>
<thead>
<tr>
<th>Type of Thyroid Cancer</th>
<th>Cancers, No. (%)</th>
<th>Size, Mean (Range), cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papillary</td>
<td>85 (81.0)</td>
<td>1.7 (1.0-7.0)</td>
</tr>
<tr>
<td>Follicular</td>
<td>Minimally invasive 6 (5.7) 3.8 (1.7-8.5)</td>
<td></td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>1 (1.0)</td>
<td>3.0a</td>
</tr>
<tr>
<td>Not otherwise specified</td>
<td>3 (2.8)</td>
<td>2.6 (1.3-5.0)</td>
</tr>
<tr>
<td>Medullary</td>
<td>6 (5.7)</td>
<td>2.6 (1.6-3.8)</td>
</tr>
<tr>
<td>Anaplastic</td>
<td>1 (1.0)</td>
<td>Extensiveb</td>
</tr>
<tr>
<td>Not specified</td>
<td>3 (2.8)</td>
<td>NA</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not available.

* No range is given for a single case.

b Involving the thyroid, invading esophagus, and carotid.

Table 1. Histologic Findings of the Study Cancers
we were unable to determine which nodule harbored cancer, all nodules were considered malignant. Nodules in patients never diagnosed as having thyroid cancer (n = 428) and benign nodules in cancer patients (n = 87) were combined to create our final control group of benign nodules (n = 515). Note that 3 cancer patients (3.1%) did not have any nodules larger than 5 mm identified on ultrasound imaging.

**Statistical Analysis**

We compared mean age, age group, sex, and year of study between patients diagnosed as having cancer and controls. We used the χ² test to determine whether the number of nodules varied by age group. We performed single-predictor modeling to assess the association between specific ultrasound imaging characteristics and cancer status using generalized estimating equations, with a compound symmetry (exchangeable) correlation structure to account for the correlated outcomes among multiple nodules within a patient. For variables that were statistically significant in the single-predictor model, we calculated diagnostic accuracy statistics (sensitivity, specificity, likelihood ratios, and predictive values).

To build the generalized estimating equation models, we added variables that were statistically significant in single-predictor models one at a time in the order of the effect size. Variables were retained if the associated P value after inclusion was <.10 for that variable. The ultrasound imaging characteristics that were retained in the final multiple-predictor model (microcalcifications, size ≥ 2 cm, and solid composition; Figure) were combined in various ways, via logic and/or criteria, to define an abnormal ultrasound imaging interpretation. The risk of cancer (predictive values) associated with each definition of an abnormal ultrasound imaging interpretation was calculated, accounting for the sampling strategy in the entire cohort. The positive predictive value (PPV) is the risk of cancer for a patient who is found to have an abnormal ultrasound imaging interpretation, and the negative predictive value is the probability of being cancer free if the ultrasound imaging result is negative for cancer. For each definition of an abnormal ultrasound imaging result, we calculated the number of cancers missed per 1000 ultrasound imaging examinations. The number of patients needed to undergo a biopsy to detect a single cancer was defined as the inverse of the PPV. We performed several sensitivity analyses to determine whether implicit assumptions in the primary analysis were reasonable. More details on the analysis are provided in the eAppendix in the Supplement.

**Results**

A total of 8806 patients underwent 11 618 thyroid ultrasound examinations during the study period, including 105 patients diagnosed as having thyroid cancer (incidence of 0.9 cancers per 100 ultrasound examinations). The cancers were diagnosed 1 day to 6.1 years after ultrasound imaging, and among control patients, there was a mean follow-up of 4.2 years (range, 2.0-10.9 years). There were no significant differences in the matching variables between cases and controls (Table 2).

Thyroid nodules were common among patients diagnosed as having thyroid cancer (96.9%) and patients not diagnosed as having thyroid cancer (56.4%) (Table 3). Among the 96 cases, 102 malignant nodules and 87 benign nodules were iden-
tified, with an increase in the number of nodules seen with advancing age. Among the 369 controls, 428 benign nodules were identified, and the number of nodules did not vary with age.

**Single-Predictor Modeling Results**

Several ultrasonographic findings were significantly associated with the odds of a nodule harboring cancer (Table 4). Microcalcifications had the strongest association with cancer; 38.2% of cancer nodules vs 5.4% of benign nodules had microcalcifications, reflecting approximately a 7-fold increase in the likelihood of cancer if microcalcifications were seen (likelihood positive ratio, 7.0; 95% CI, 6.0-8.2) and a 30% reduction in the likelihood of cancer if microcalcifications were not seen (likelihood negative ratio, 0.65; 95% CI, 0.56-0.76). The corresponding odds ratio (OR) was 11.6 (95% CI, 6.5-20.0).

Coarse calcifications, nodule composition, nodule echogenicity, central vascularity, margins, and shape were also each significantly associated with cancer, but the magnitude of association was smaller, with ORs ranging from 1.6 to 2.9. Rim calcifications and comet tail artifacts, peripheral vascularity, and the presence of a halo were not associated with the likelihood of cancer.

The oddsof cancer increased with nodulesize, and the largest nodules had the greatest odds of cancer (likelihood ratio, 1.8; 95% CI, 1.5-2.1; and OR, 3.1; 95% CI, 1.8-5.2) for nodules larger than 2 cm compared with nodules smaller than 1 cm. Simple cysts never reflected cancer.

**Multiple-Predictor Nodule Modeling Results**

Only 3 nodule characteristics were significantly associated with the risk of cancer in the multiple-predictor modeling: micro-
calcifications (OR, 8.1; 95% CI, 3.8-17.3; P < .001), size greater than 2 cm (OR, 3.6; 95% CI, 1.7-7.6; P = .001), and an entirely solid composition (OR, 4.0; 95% CI, 1.7-9.2; P = .001). The nodule characteristics were not significantly associated with the risk of cancer, and including them in the definition of an abnormal nodule added less than 2% cancer detection.

The accuracy of the several definitions of an abnormal ultrasound imaging interpretation is given in Table 5. If any 1 of the 3 characteristics is used to prompt biopsy, most cases of thyroid cancer would be detected (sensitivity, 0.88; 95% CI, 0.80-0.94) at a false-positive rate of 0.44 (95% CI, 0.43-0.45). The high false-positive rate of this approach is reflected in a low PPV (ie, risk of cancer) of 1.8% (95% CI, 1.5%-2.2%) when a single characteristic is used to prompt biopsy, and 56 biopsies will be required per cancer diagnosed. If 2 abnormal ultrasound imaging characteristics were required to prompt biopsy, the sensitivity and false-positive rates would be lower (sensitivity, 0.52; 95% CI, 0.42-0.62; false-positive rate, 0.07; 95% CI, 0.07-0.08), and the risk of cancer in those with an ultrasound image suggestive of cancer would be higher (PPV, 6.2%; 95% CI, 4.7%-8.7%) and fewer biopsies (n = 16) would be required per cancer diagnosed. Compared with existing guidelines that recommend biopsy of all thyroid nodules greater than 5 mm, adoption of this more stringent rule requiring 2 abnormal characteristics to prompt biopsy would reduce unnecessary biopsies by 90% while maintaining a low risk of cancer in patients in whom biopsy is deferred (ie, 5 cancers [0.5%] per 1000 ultrasound examinations).

The most specific definition of an abnormal ultrasound image is one requiring all 3 abnormal characteristics to prompt biopsy; however, this definition would detect only a small proportion of cancers (sensitivity, 0.07; 95% CI, 0.03-0.14) but would have a high positive likelihood ratio of 28 (95% CI, 23-34).

The tradeoff between the different definitions of an abnormal ultrasound imaging result and test accuracy is shown in the Figure. As the number of criteria required to prompt biopsy increases, the number of missed cancers (false-negative results) increases, and the number of patients who undergo biopsy to detect a cancer will decrease. For example, if 2 criteria instead of 1 are required to prompt biopsy, the rate of missed cancers among patients who do not undergo biopsy increases from 2 to 5 per 1000 ultrasound examinations, whereas the number of biopsies needed to detect a cancer decreases from 56 to 16.

**Predictive Values: Patient-Level Analysis**

The risk of cancer based on the appearance of the thyroid on the ultrasound image is given in Table 6. The risk of cancer is low among patients with a homogeneous thyroid, in which no nodules were identified (0.63 cancers per 1000 patients). The risk of cancer is also low in patients in whom the only ultrasound imaging characteristic is a simple cyst (0.32 cancers per 1000 patients).

If the presence of a single abnormal characteristic is used to define an abnormal ultrasound examination result, patients with a normal examination result will have a risk of can-

---

**Table 4. Accuracy of Individual Ultrasound Imaging Characteristics**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cases, No. (%) (n = 102)</th>
<th>Controls, No. (%) (n = 515)</th>
<th>Likelihood Ratio (95% CI)</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microcalcifications</td>
<td>39 (38.2)</td>
<td>28 (5.4)</td>
<td>7.0 (6.0-8.2)</td>
<td>0.65 (0.56-0.76)</td>
</tr>
<tr>
<td>Coarse calcifications</td>
<td>13 (12.7)</td>
<td>34 (6.6)</td>
<td>1.9 (1.6-2.3)</td>
<td>0.93 (0.78-1.1)</td>
</tr>
<tr>
<td>Rim calcifications</td>
<td>5 (4.9)</td>
<td>23 (4.5)</td>
<td>1.1 (0.91-1.3)</td>
<td>1.0 (0.82-1.2)</td>
</tr>
<tr>
<td>Comet tail artifact</td>
<td>17 (16.7)</td>
<td>77 (15.0)</td>
<td>1.1 (0.93-1.3)</td>
<td>0.98 (0.82-1.2)</td>
</tr>
<tr>
<td>Composition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entirely solid</td>
<td>68 (66.7)</td>
<td>220 (42.7)</td>
<td>1.6 (1.4-1.8)</td>
<td>0.58 (0.51-0.66)</td>
</tr>
<tr>
<td>Entirely cystic</td>
<td>0</td>
<td>37 (7.2)</td>
<td>0.034 (0.028-0.041)</td>
<td>1.1 (0.88-1.3)</td>
</tr>
<tr>
<td>Mixed solid and cystic</td>
<td>34 (33.3)</td>
<td>248 (48.2)</td>
<td>0.69 (0.58-0.82)</td>
<td>1.3 (1.1-1.5)</td>
</tr>
<tr>
<td>Echogenicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypoechoic to strap</td>
<td>16 (15.7)</td>
<td>33 (6.4)</td>
<td>2.4 (2.0-2.9)</td>
<td>0.89 (0.75-1.1)</td>
</tr>
<tr>
<td>Isoechoic or hyperechoic to strap</td>
<td>51 (50.0)</td>
<td>198 (38.4)</td>
<td>1.3 (1.1-1.5)</td>
<td>0.81 (0.70-0.94)</td>
</tr>
<tr>
<td>Isoechoic to thyroid</td>
<td>30 (29.4)</td>
<td>209 (40.6)</td>
<td>0.72 (0.61-0.86)</td>
<td>1.2 (1.0-1.4)</td>
</tr>
<tr>
<td>Hyperechoic to thyroid</td>
<td>4 (3.9)</td>
<td>38 (7.4)</td>
<td>0.53 (0.44-0.64)</td>
<td>1.0 (0.86-1.3)</td>
</tr>
<tr>
<td>Central vascular flow</td>
<td>40 (39.2)</td>
<td>136 (26.4)</td>
<td>1.5 (1.3-1.7)</td>
<td>0.83 (0.71-0.97)</td>
</tr>
<tr>
<td>Peripheral vascular flow</td>
<td>49 (48.0)</td>
<td>201 (39.0)</td>
<td>1.2 (1.1-1.4)</td>
<td>0.85 (0.73-0.99)</td>
</tr>
<tr>
<td>Halo</td>
<td>28 (27.4)</td>
<td>116 (22.5)</td>
<td>1.2 (1.0-1.4)</td>
<td>0.94 (0.79-1.1)</td>
</tr>
<tr>
<td>Margins (ill defined or lobulated)</td>
<td>61 (59.8)</td>
<td>212 (41.2)</td>
<td>1.5 (1.3-1.7)</td>
<td>0.68 (0.60-0.78)</td>
</tr>
<tr>
<td>Shape (smaller than wide)</td>
<td>18 (17.6)</td>
<td>42 (8.2)</td>
<td>2.2 (1.8-2.6)</td>
<td>0.89 (0.75-1.1)</td>
</tr>
<tr>
<td>Size, cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤1</td>
<td>30 (29.4)</td>
<td>248 (48.2)</td>
<td>0.61 (0.51-0.73)</td>
<td>1.4</td>
</tr>
<tr>
<td>1.01-2</td>
<td>38 (37.2)</td>
<td>169 (32.8)</td>
<td>1.1 (0.97-1.3)</td>
<td>0.94</td>
</tr>
<tr>
<td>&gt;2</td>
<td>34 (33.3)</td>
<td>97 (18.8)</td>
<td>1.8 (1.5-2.1)</td>
<td>0.82</td>
</tr>
</tbody>
</table>

* Each characteristic was considered individually and was not adjusted for the presence of the other findings. The odds ratios were adjusted for multiple nodules within study participants, whereas the accuracy statistics were not adjusted for multiple nodules within study participants. The numbers may not sum because nodules in which the ultrasound imaging characteristic was missing are included in the calculations of sensitivity and specificity but are not listed in the table.

* P < .01.

* P < .05.

* Entirely cystic includes cystic honeycomb. To calculate the likelihood ratio, 0.25 was added to zero cells. Because entirely cystic nodules had no cancer cases, an odds ratio could not be calculated.
Thyroid nodules are common, most (98.4%) are benign, highlighting the importance of being prudent in deciding which nodules should be sampled to reduce unnecessary biopsies. Unnecessary tissue sampling not only is invasive and costly but also leads to repeated sampling and unnecessary open surgical procedures because up to one-third of fine-needle aspiration biopsies may be nondiagnostic, requiring open surgical biopsy for diagnosis. We found that only 3 ultrasound imaging characteristics—microcalcifications, size larger than 2 cm, and entirely solid composition—were statistically significantly associated with the risk of cancer and that, when used in combination, these 3 characteristics could be used to help determine which nodules should be sampled. Simple cysts are essentially never malignant and should not be sampled.

There are many ways to characterize the accuracy of ultrasound imaging. We believe the risk of cancer (PPV) is the most relevant to patients and physicians, and ours is the first study, to our knowledge, that permits estimating this risk. A patient’s risk of harboring cancer ranges from 2 per 1000 patients if a patient has a nodule with none of the 3 characteristics identified, 62 per 1000 patients if a patient has a nodule with a single characteristic, 62 per 1000 patients if a patient has a nodule with 2 abnormal characteristics, and 960 per 1000 patients if a patient has a nodule with all 3 characteristics. Although there is growing concern regarding overdiagnosis and overtreatment across all areas of medicine, there are no well-established guidelines regarding what risk is low enough that an imaging finding can be ignored. In other areas of diagnostic testing, for example, when assessing patients at risk for acute coronary syndrome, many withhold surgery to minimize morbidity and mortality, a risk of less than 1% or 0.5% is often considered sufficiently low that further evaluation is deemed unnecessary.

Discussion
Thyroid nodules are extremely common. Even among patients selected as controls in our study, 56.4% had thyroid nodules greater than 5 mm, and nearly one-third had multiple nodules. In contrast to previous reports that have suggested the prevalence of cancer in thyroid nodules as high as 23%, we found that only 1.6% of patients who had 1 or more thyroid nodules 5 mm or greater harbored cancer. Thus, although thyroid nodules are common, most (98.4%) are benign, highlighting the importance of being prudent in deciding which nodules should be sampled to reduce unnecessary biopsies. Unnecessary tissue sampling not only is invasive and costly but also leads to repeated sampling and unnecessary open surgical procedures because up to one-third of fine-needle aspiration biopsies may be nondiagnostic, requiring open surgical biopsy for diagnosis. We found that only 3 ultrasound imaging characteristics—microcalcifications, size larger than 2 cm, and entirely solid composition—were statistically significantly associated with the risk of cancer and that, when used in combination, these 3 characteristics could be used to help determine which nodules should be sampled. Simple cysts are essentially never malignant and should not be sampled.

There are many ways to characterize the accuracy of ultrasound imaging. We believe the risk of cancer (PPV) is the most relevant to patients and physicians, and ours is the first study, to our knowledge, that permits estimating this risk. A patient’s risk of harboring cancer ranges from 2 per 1000 patients if a patient has a nodule with none of the 3 characteristics identified, 62 per 1000 patients if a patient has a nodule with a single characteristic, 62 per 1000 patients if a patient has a nodule with 2 abnormal characteristics, and 960 per 1000 patients if a patient has a nodule with all 3 characteristics. Although there is growing concern regarding overdiagnosis and overtreatment across all areas of medicine, there are no well-established guidelines regarding what risk is low enough that an imaging finding can be ignored. In other areas of diagnostic testing, for example, when assessing patients at risk for acute coronary syndrome, many withhold surgery to minimize morbidity and mortality, a risk of less than 1% or 0.5% is often considered sufficiently low that further evaluation is deemed unnecessary.

Discussion
Thyroid nodules are extremely common. Even among patients selected as controls in our study, 56.4% had thyroid nodules greater than 5 mm, and nearly one-third had multiple nodules. In contrast to previous reports that have suggested the prevalence of cancer in thyroid nodules as high as 23%, we found that only 1.6% of patients who had 1 or more thyroid nodules 5 mm or greater harbored cancer. Thus, although thyroid nodules are common, most (98.4%) are benign, highlighting the importance of being prudent in deciding which nodules should be sampled to reduce unnecessary biopsies. Unnecessary tissue sampling not only is invasive and costly but also leads to repeated sampling and unnecessary open surgical procedures because up to one-third of fine-needle aspiration biopsies may be nondiagnostic, requiring open surgical biopsy for diagnosis. We found that only 3 ultrasound imaging characteristics—microcalcifications, size larger than 2 cm, and entirely solid composition—were statistically significantly associated with the risk of cancer and that, when used in combination, these 3 characteristics could be used to help determine which nodules should be sampled. Simple cysts are essentially never malignant and should not be sampled.

There are many ways to characterize the accuracy of ultrasound imaging. We believe the risk of cancer (PPV) is the most relevant to patients and physicians, and ours is the first study, to our knowledge, that permits estimating this risk. A patient’s risk of harboring cancer ranges from 2 per 1000 patients if a patient has a nodule with none of the 3 characteristics identified, 62 per 1000 patients if a patient has a nodule with a single characteristic, 62 per 1000 patients if a patient has a nodule with 2 abnormal characteristics, and 960 per 1000 patients if a patient has a nodule with all 3 characteristics. Although there is growing concern regarding overdiagnosis and overtreatment across all areas of medicine, there are no well-established guidelines regarding what risk is low enough that an imaging finding can be ignored. In other areas of diagnostic testing, for example, when assessing patients at risk for acute coronary syndrome, many withhold surgery to minimize morbidity and mortality, a risk of less than 1% or 0.5% is often considered sufficiently low that further evaluation is deemed unnecessary.

Discussion
Thyroid nodules are extremely common. Even among patients selected as controls in our study, 56.4% had thyroid nodules greater than 5 mm, and nearly one-third had multiple nodules. In contrast to previous reports that have suggested the prevalence of cancer in thyroid nodules as high as 23%, we found that only 1.6% of patients who had 1 or more thyroid nodules 5 mm or greater harbored cancer. Thus, although thyroid nodules are common, most (98.4%) are benign, highlighting the importance of being prudent in deciding which nodules should be sampled to reduce unnecessary biopsies. Unnecessary tissue sampling not only is invasive and costly but also leads to repeated sampling and unnecessary open surgical procedures because up to one-third of fine-needle aspiration biopsies may be nondiagnostic, requiring open surgical biopsy for diagnosis. We found that only 3 ultrasound imaging characteristics—microcalcifications, size larger than 2 cm, and entirely solid composition—were statistically significantly associated with the risk of cancer and that, when used in combination, these 3 characteristics could be used to help determine which nodules should be sampled. Simple cysts are essentially never malignant and should not be sampled.

There are many ways to characterize the accuracy of ultrasound imaging. We believe the risk of cancer (PPV) is the most relevant to patients and physicians, and ours is the first study, to our knowledge, that permits estimating this risk. A patient’s risk of harboring cancer ranges from 2 per 1000 patients if a patient has a nodule with none of the 3 characteristics identified, 62 per 1000 patients if a patient has a nodule with a single characteristic, 62 per 1000 patients if a patient has a nodule with 2 abnormal characteristics, and 960 per 1000 patients if a patient has a nodule with all 3 characteristics. Although there is growing concern regarding overdiagnosis and overtreatment across all areas of medicine, there are no well-established guidelines regarding what risk is low enough that an imaging finding can be ignored. In other areas of diagnostic testing, for example, when assessing patients at risk for acute coronary syndrome, many withhold surgery to minimize morbidity and mortality, a risk of less than 1% or 0.5% is often considered sufficiently low that further evaluation is deemed unnecessary.

Discussion
Thyroid nodules are extremely common. Even among patients selected as controls in our study, 56.4% had thyroid nodules greater than 5 mm, and nearly one-third had multiple nodules. In contrast to previous reports that have suggested the prevalence of cancer in thyroid nodules as high as 23%, we found that only 1.6% of patients who had 1 or more thyroid nodules 5 mm or greater harbored cancer. Thus, although thyroid nodules are common, most (98.4%) are benign, highlighting the importance of being prudent in deciding which nodules should be sampled to reduce unnecessary biopsies. Unnecessary tissue sampling not only is invasive and costly but also leads to repeated sampling and unnecessary open surgical procedures because up to one-third of fine-needle aspiration biopsies may be nondiagnostic, requiring open surgical biopsy for diagnosis. We found that only 3 ultrasound imaging characteristics—microcalcifications, size larger than 2 cm, and entirely solid composition—were statistically significantly associated with the risk of cancer and that, when used in combination, these 3 characteristics could be used to help determine which nodules should be sampled. Simple cysts are essentially never malignant and should not be sampled.

There are many ways to characterize the accuracy of ultrasound imaging. We believe the risk of cancer (PPV) is the most relevant to patients and physicians, and ours is the first study, to our knowledge, that permits estimating this risk. A patient’s risk of harboring cancer ranges from 2 per 1000 patients if a patient has a nodule with none of the 3 characteristics identified, 62 per 1000 patients if a patient has a nodule with a single characteristic, 62 per 1000 patients if a patient has a nodule with 2 abnormal characteristics, and 960 per 1000 patients if a patient has a nodule with all 3 characteristics. Although there is growing concern regarding overdiagnosis and overtreatment across all areas of medicine, there are no well-established guidelines regarding what risk is low enough that an imaging finding can be ignored. In other areas of diagnostic testing, for example, when assessing patients at risk for acute coronary syndrome, many withhold surgery to minimize morbidity and mortality, a risk of less than 1% or 0.5% is often considered sufficiently low that further evaluation is deemed unnecessary.

Discussion
Thyroid nodules are extremely common. Even among patients selected as controls in our study, 56.4% had thyroid nodules greater than 5 mm, and nearly one-third had multiple nodules. In contrast to previous reports that have suggested the prevalence of cancer in thyroid nodules as high as 23%, we found that only 1.6% of patients who had 1 or more thyroid nodules 5 mm or greater harbored cancer. Thus, although thyroid nodules are common, most (98.4%) are benign, highlighting the importance of being prudent in deciding which nodules should be sampled to reduce unnecessary biopsies. Unnecessary tissue sampling not only is invasive and costly but also leads to repeated sampling and unnecessary open surgical procedures because up to one-third of fine-needle aspiration biopsies may be nondiagnostic, requiring open surgical biopsy for diagnosis. We found that only 3 ultrasound imaging characteristics—microcalcifications, size larger than 2 cm, and entirely solid composition—were statistically significantly associated with the risk of cancer and that, when used in combination, these 3 characteristics could be used to help determine which nodules should be sampled. Simple cysts are essentially never malignant and should not be sampled.

There are many ways to characterize the accuracy of ultrasound imaging. We believe the risk of cancer (PPV) is the most relevant to patients and physicians, and ours is the first study, to our knowledge, that permits estimating this risk. A patient’s risk of harboring cancer ranges from 2 per 1000 patients if a patient has a nodule with none of the 3 characteristics identified, 62 per 1000 patients if a patient has a nodule with a single characteristic, 62 per 1000 patients if a patient has a nodule with 2 abnormal characteristics, and 960 per 1000 patients if a patient has a nodule with all 3 characteristics. Although there is growing concern regarding overdiagnosis and overtreatment across all areas of medicine, there are no well-established guidelines regarding what risk is low enough that an imaging finding can be ignored. In other areas of diagnostic testing, for example, when assessing patients at risk for acute coronary syndrome, many withhold surgery to minimize morbidity and mortality, a risk of less than 1% or 0.5% is often considered sufficiently low that further evaluation is deemed unnecessary.
immediate treatment. Thus, given the favorable prognosis of most thyroid cancers even without treatment, a risk of cancer of 0.5% among those with a negative examination is unlikely to prove beneficial given that our results ascribe to the theoretical metastatic cancer to the thyroid beyond the link of the cohort with data from a comprehensive cancer registry, which allows accurate assessment of the true underlying prevalence of cancer. The analysis has several limitations. We did not have accurate information about why patients underwent imaging, and the risk of cancer may vary based on the reasons patients received sonograms. We did not stratify the results by the histologic type of cancer, although most included cancers were papillary cancer, as is the case for thyroid cancer in general. There are several ultrasonographic features that we did not assess, but these are rare, such as extracapsular growth or abnormal lymph nodes. We did not include the theoretical metastatic cancer to the thyroid because these would not be captured in the cancer registry data. However, we also linked to the local pathology database, and no cases of metastatic cancer were identified.

The increased use and improved technical quality of ultrasonographic imaging has given rise to the detection of multiple morphologic characteristics, without clear criteria for what nodules need further evaluation, resulting in greater tissue sampling and excessive treatment. In mammography, the adoption of uniform interpretation standards through the Breast Imaging Reporting and Data System has been useful in allowing comparative effectiveness work in breast imaging and efforts to standardize the interpretation of mammograms. Similar adoption of uniform standards for the interpretation of thyroid sonograms would be a first step toward standardizing the diagnosis and treatment of thyroid cancer and limiting unnecessary diagnostic testing and treatment.

Our study was designed to determine how to reduce unnecessary and excessive thyroid surveillance and biopsy. Our study does not provide evidence as to whether the detection of thyroid cancers will lead to improved patient outcomes. There has been a recent increase in the observed incidence of small and microthyroid cancer without a corresponding change in the thyroid cancer mortality rate, raising the question of whether there is benefit to the earlier diagnosis or treatment of incidental thyroid cancer.

Many previous studies have assessed the risk of cancer associated with the appearance of the thyroid on ultrasound imaging. All previous studies have inflated the association between nodule characteristics and cancer risk because they limited their analysis to nodules that underwent biopsy. For example, Ahn et al compared various existing guidelines for prompting fine-needle aspiration in a sample of 1398 patients who had undergone biopsy. In this sample, 20% of the included patients had cancer, contrasting with the 1.5% cancer rate in our study. They report that the PPV for cancer in a patient who has microcalcifications is 85.1%, whereas using our population-based approach without ascertainment bias, we found a PPV of 5.8%. We considered many nodule characteristics endorsed by other authors, but when put into the multiple-predictor models, most of the characteristics were not significantly associated with cancer risk.

It is widely reported that the number of benign thyroid nodules increases with age. We observed this relationship among patients diagnosed as having cancer but not among patients without cancer.

The main strength of our study is the large sample size and the linkage of the cohort with data from a comprehensive cancer registry, which allows accurate assessment of the true underlying prevalence of cancer. The analysis has several limitations. We did not have accurate information about why patients underwent imaging, and the risk of cancer may vary based on the reasons patients received sonograms. We did not stratify the results by the histologic type of cancer, although most included cancers were papillary cancer, as is the case for thyroid cancer in general. There are several ultrasonographic features that we did not assess, but these are rare, such as extracapsular growth or abnormal lymph nodes. We did not include the theoretical metastatic cancer to the thyroid because these would not be captured in the cancer registry data. However, we also linked to the local pathology database, and no cases of metastatic cancer were identified.

The increased use and improved technical quality of ultrasonographic imaging has given rise to the detection of multiple morphologic characteristics, without clear criteria for what nodules need further evaluation, resulting in greater tissue sampling and excessive treatment. In mammography, the adoption of uniform interpretation standards through the Breast Imaging Reporting and Data System has been useful in allowing comparative effectiveness work in breast imaging and efforts to standardize the interpretation of mammograms. Similar adoption of uniform standards for the interpretation of thyroid sonograms would be a first step toward standardizing the diagnosis and treatment of thyroid cancer and limiting unnecessary diagnostic testing and treatment.

### Table 6. Risk of Thyroid Cancer Based on the Appearance of the Thyroid Gland and the Characteristics of Any Nodules Identified

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. of Cancers per 1000 Patients (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest risk</td>
<td></td>
</tr>
<tr>
<td>Thyroid has only nodule(s) that are entirely cystic</td>
<td>0.32 (0.1-1.8)</td>
</tr>
<tr>
<td>Homogenous thyroid</td>
<td>0.63 (0.1-1.9)</td>
</tr>
<tr>
<td>Very low</td>
<td></td>
</tr>
<tr>
<td>Thyroid has no solid nodule(s), no nodule &gt;2 cm, and no nodule with microcalcification</td>
<td>2 (1-3)</td>
</tr>
<tr>
<td>Thyroid has no nodule(s) with 2 or more suspicious characteristic</td>
<td>5 (4-6)</td>
</tr>
<tr>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Thyroid has at least one nodule that is entirely solid or &gt;2 cm or has microcalcification</td>
<td>18 (15-22)</td>
</tr>
<tr>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Thyroid has at least 1 nodule with 2 suggestive characteristics</td>
<td>62 (47-80)</td>
</tr>
<tr>
<td>Thyroid has at least 1 nodule with microcalcification</td>
<td>82 (59-110)</td>
</tr>
<tr>
<td>Thyroid has at least 1 nodule with microcalcification or is both solid and &gt;2 cm</td>
<td>58 (44-75)</td>
</tr>
<tr>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Thyroid has at least 1 entirely solid nodule with microcalcification and a size &gt;2 cm</td>
<td>960 (803-991)</td>
</tr>
</tbody>
</table>

unnecessary. If a thyroid cancer risk less than 0.5% is considered acceptable for those in whom biopsy is deferred, using microcalcifications or the combined observations of a large (>2-cm) solid nodule as the only features to prompt biopsy reflects a good choice. In comparison with various guidelines that recommend biopsies in a larger number of patients, limiting biopsy to nodules that fulfill this definition would reduce the number of biopsies by as much as 90% while maintaining a low cancer rate of 5 per 1000 patients among individuals who do not undergo thyroid sampling. Most thyroid cancers have a favorable prognosis, with a 20-year survival greater than 97% seen even among patients who do not receive immediate treatment. Thus, given the favorable prognosis of most thyroid cancers even without treatment, a risk of cancer of 0.5% among those with a negative examination result seems to balance between detection and unnecessary tissue sampling. Ongoing ultrasound imaging surveillance of patients with nodules who do not meet the criteria for biopsy is unlikely to prove beneficial given that our results ascribe these patients a low risk of cancer for as long as 10 years after imaging.

Our study was designed to determine how to reduce unnecessary and excessive thyroid surveillance and biopsy. Our study does not provide evidence as to whether the detection of thyroid cancers will lead to improved patient outcomes. There has been a recent increase in the observed incidence of small and microthyroid cancer without a corresponding change in the thyroid cancer mortality rate, raising the question of whether there is benefit to the earlier diagnosis or treatment of incidental thyroid cancer. Many previous studies have assessed the risk of cancer associated with the appearance of the thyroid on ultrasound imaging. All previous studies have inflated the association between nodule characteristics and cancer risk because they limited their analysis to nodules that underwent biopsy. For example, Ahn et al compared various existing guidelines for prompting fine-needle aspiration in a sample of 1398 patients who had undergone biopsy. In this sample, 20% of the included patients had cancer, contrasting with the 1.5% cancer rate in our study. They report that the PPV for cancer in a patient who has microcalcifications is 85.1%, whereas using our population-based approach without ascertainment bias, we found a PPV of 5.8%. We considered many nodule characteristics endorsed by other authors, but when put into the multiple-predictor models, most of the characteristics were not significantly associated with cancer risk.

It is widely reported that the number of benign thyroid nodules increases with age. We observed this relationship among patients diagnosed as having cancer but not among patients without cancer.

The main strength of our study is the large sample size and the linkage of the cohort with data from a comprehensive cancer registry, which allows accurate assessment of the true underlying prevalence of cancer. The analysis has several limitations. We did not have accurate information about why patients underwent imaging, and the risk of cancer may vary based on the reasons patients received sonograms. We did not stratify the results by the histologic type of cancer, although most included cancers were papillary cancer, as is the case for thyroid cancer in general. There are several ultrasonographic features that we did not assess, but these are rare, such as extracapsular growth or abnormal lymph nodes. We did not include the theoretical metastatic cancer to the thyroid because these would not be captured in the cancer registry data. However, we also linked to the local pathology database, and no cases of metastatic cancer were identified.

The increased use and improved technical quality of ultrasonographic imaging has given rise to the detection of multiple morphologic characteristics, without clear criteria for what nodules need further evaluation, resulting in greater tissue sampling and excessive treatment. In mammography, the adoption of uniform interpretation standards through the Breast Imaging Reporting and Data System has been useful in allowing comparative effectiveness work in breast imaging and efforts to standardize the interpretation of mammograms. Similar adoption of uniform standards for the interpretation of thyroid sonograms would be a first step toward standardizing the diagnosis and treatment of thyroid cancer and limiting unnecessary diagnostic testing and treatment.
Thyroid Cancer Risk

ORIGINAL INVESTIGATION

Thyroid Cancer Risk

REFERENCES


33. Grady D, Redberg RF. Less is more: how less health care can result in better health. Arch Intern Med. 2010;170(9):749-750.


35. Davies L, Ouellette M, Hunter M, Welch HG. The increasing incidence of small thyroid cancers: where are the cases coming from? Laryngoscope. 2010;120(12):2446-2451.

The Importance, and Important Limitations, of Ultrasound Imaging for Evaluating Thyroid Nodules

Erik K. Alexander, MD; David Cooper, MD

The care of a patient with a thyroid nodule has become increasingly complex. Although early detection and intervention are sought for most human malignant tumors, these mandates have proved less certain for thyroid cancer because many affected patients (even when untreated) do not die of their disease. Thus, physicians increasingly seek noninvasive means by which to identify patients who can avoid unnecessary diagnostic or therapeutic intervention. During the last 2 decades, sonographic imaging of the thyroid has become the most valued tool. The investigation published in this week's *JAMA Internal Medicine* by Smith-Bindman and colleagues lends further support. The authors performed a large-scale analysis of all thyroid ultrasonographic examinations performed at a single institution during a 5-year period. Radiologic findings were linked with the California Cancer Registry, allowing masked assessment of sonographic features that convey benign or malignant disease. This investigation confirms the importance of analyzing parenchymal tissue (solid or cystic), measuring nodule size, and identifying sonographic features, such as microcalcifications, irregular margins, and increased vascular flow. These results are important because the large-scale and masked analysis by Smith-Bindman et al validates results from lessstringently performed, retrospective analyses. However, the data about suspicious nodule characteristics were derived from a total of only 96 malignant tumors.

Accurate interpretation and effective translation of these findings into clinical practice is critical. Currently, expert guidelines generally recommend diagnostic fine-needle aspiration biopsy (FNA) only for thyroid nodules larger than 1.0 to 1.5 cm in maximal diameter. This is because data confirm the indolent, nonharmful nature of most thyroid microcarcinomas smaller than 1 cm in diameter. Furthermore, FNA cytologic diagnoses are nonspecific, often conveying indeterminate results in which cytologic abnormalities raise concern for malignancy, although most are ultimately found to be benign. In such cases, the risk of cancer must also be balanced against the risk of unnecessary surgical intervention. In the current study, the retrospective, case-control design mandates readers carefully compare cohort characteristics and diagnostic processes to expected standards. In doing so, some features of this investigation are notable. For example, it has been well established that the prevalence of thyroid cancer in a population of consecutively evaluated nodules approximates 8% to 15%. The current study, however, analyzed a cohort in which cancer prevalence was only 1.6%, whereas 29% to 48% of nodules were smaller than 1 cm. In addition, the number of nodules in the thyroid increased with age in patients with cancer but not among controls. These facts suggest a nonrepresentative study population and an increased potential for sampling bias.

It also appears unlikely that all euthyroid patients with thyroid nodules were evaluated with ultrasonography-guided FNA. Because this would be necessary to fully understand the tissue characteristics (benign or malignant) of each nodule, the author's assumption that all patients not listed in the cancer registry have benign disease is unfounded. Thus, notwithstanding the high-quality nature of the author's sonographic review and data analysis, the study's epidemiologic interpretations are not readily transferable to clinical practice. Specifically, test metrics, such as positive or negative predictive value and calculations of number needed to diagnose, are influenced by population characteristics and the prevalence of disease.

Does this imply that the author's conceptual framework should not be applied to thyroid nodule evaluation? The answer is most certainly no. Sonographic features are, and should remain, the principal tool to guide thyroid cancer risk assessment. Solid nodules larger than 1 cm, especially with microcalcifications, should be the primary focus of diagnostic evaluation. Conversely, purely cystic nodules can be safely monitored. Unfortunately, the authors also do not address the spongiform nodule, which also has a high benignity rate. Although they combine cystic honeycomb nodules with purely cystic nodules, it is unclear that such nodules are the same as spongiform nodules. All sonographic findings should be considered so that nodules can be prioritized for possible tissue evaluation. Similarly, to allow clinicians to use such sonographic characteristics in clinical decision making, reporting by radiologists must also include as many descriptors as possible. All too often, ultrasonographic reports provide nodule...